ÖREBRO
UNIVERSITET

Lösning till tentamen i

Matematisk statistik och sannolikhetslära

MA506G

2021-01-02

---

1. Inför en muntlig tentamen har studenterna fått ut nio uppgifter som de ska förbereda svar på. Av dessa kommer läraren för varje student att välja ut en uppgift. För att få variation väljer läraren inte samma uppgift till två på varandra följande studenter, men i övrigt sker valet helt slumpmässigt och likafördelat.

   De tre första studenterna heter, i ordning, Anna, Basam och Cecilia. Trots idoga studier vet Anna och Basam inte hur de ska lösa uppgifterna 8 och 9, och Cecilia klarar inte 6, 7 eller 8, men övriga uppgifter är de säkra på.

   (a) Beräkna sannolikheten att både Anna och Basam klarar muntan. [2p]

   (b) Beräkna sannolikheten att Cecilia klarar muntan om Anna och Basam har klarat den. [4p]

   (c) Efter muntan samlas studenterna på tentapub och Cecilia berättar för Diana att hon klarat muntan. De vet dock inte hur det gått för Anna och Basam. Givet denna information, beräkna sannolikheten att Anna och Basam har klarat muntan. [4p]

   Lösning:

   (a) Låt händelserna att Anna. Basam och Cecilia klarar muntan betecknas A, B och C. vi får då

   $$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B|A) = \frac{7}{9} \cdot \frac{6}{8} = \frac{7}{12}.$$

   (b) Eftersom Anna och Basam klarade muntan så fick Basam inte uppgift 8 eller 9. Sannolikheten att Cecilia klarar muntan beror på om Basams uppgift var någon av uppgifterna 1-5 eller om han fick en av uppgifterna 6-7.

Låt Basams uppgift betecknas u. Med hjälp av lagen om total sannolikhet får vi

$$\begin{aligned}\mathbf{P}(C|A\cap B) &= \mathbf{P}(u\le 5)\mathbf{P}(C|u\le 5)+\mathbf{P}(u\ge 6)\mathbf{P}(C|u\ge 6)\\ &= \frac{5}{7}\cdot\frac{5}{8}+\frac{2}{7}\cdot\frac{6}{8}=\frac{25+12}{56}=\frac{37}{56}.\end{aligned}$$

(c) Med hjälp av Bayes sats får vi

$$\mathbf{P}(A\cap B|C)=\frac{\mathbf{P}(A\cap B)\mathbf{P}(C|A\cap B)}{\mathbf{P}(C)}=\frac{\frac{7}{12}\cdot\frac{37}{56}}{\frac{2}{3}}=\frac{37}{96}\cdot\frac{3}{2}=\frac{37}{64}.$$

2. Slumpvariablerna X och Y har båda täthetsfunktionen

$$f_X(x)=\begin{cases} cxe^x & 0\le x\le 1;\\ 0 & \text{annars.}\end{cases}$$

(a) Beräkna konstanten c. [2p]

(b) Beräkna $\mathbf{E}(X)$ och $\mathbf{V}(X)$. [4p]

(c) Beräkna $\mathbf{E}(2X-Y)$ och $\mathbf{V}(2X-Y)$. [4p]

Lösning:

(a) Vi har, med hjälp av partialintegrering,

$$\begin{aligned}1 &= \int_0^1 f_X(x)\,\mathrm{d}x\\ &= c\int_0^1 xe^x\,\mathrm{d}x\\ &= c\left[xe^x\right]_0^1 - c\int_0^1 e^x\,\mathrm{d}x\\ &= ce - c\left[e^x\right]_0^1\\ &= ce - c(e-1) = c.\end{aligned}$$

(b) Med hjälp av beräkningen ovan och ytterligare partialintegreringar får vi

$$\begin{aligned}\mathbf{E}(X) &= \int_0^1 xf_X(x)\,\mathrm{d}x\\ &= \int_0^1 x^2e^x\,\mathrm{d}x\\ &= \left[x^2e^x\right]_0^1 - \int_0^1 2xe^x\,\mathrm{d}x\\ &= e - 2\int_0^1 xe^x\,\mathrm{d}x = e-2,\end{aligned}$$

och

$$\begin{aligned}\mathbf{E}(X^2) &= \int_0^1 x^2 f_X(x)\,\mathrm{d}x\\ &= \int_0^1 x^3\mathrm{e}^x\,\mathrm{d}x\\ &= \left[x^3\mathrm{e}^x\right]_0^1 - \int_0^1 3x^2\mathrm{e}^x\,\mathrm{d}x\\ &= \mathrm{e} - 3(\mathrm{e}-2) = 6-2\mathrm{e},\end{aligned}$$

som tillsammans även ger

$$V(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = 6 - 2\mathrm{e} - (\mathrm{e}-2)^2 = 2 + 2\mathrm{e} - \mathrm{e}^2.$$

(c) Eftersom X och Y har samma täthetsfunktion har de också samma väntevärde och varians. Enligt räknelagarna för väntevärden och varianser har vi, under antagandet att X och Y är oberoende (vilket borde ha stått i uppgiftsformuleringen), att

$$\mathbf{E}(2X-Y) = 2\mathbf{E}(X) - \mathbf{E}(Y) = 2(\mathrm{e}-2) - (\mathrm{e}-2) = \mathrm{e}-2$$

och

$$\mathbf{V}(2X-Y) = 2^2\mathbf{V}(X) + (-1)^2\mathbf{V}(Y) = 5(2+2\mathrm{e}-\mathrm{e}^2).$$

3. I en datorkrets sitter 3 transistorer som har livslängderna $X_k \sim \mathrm{Exp}(10^{-4})$ för $k \in \{1,2,3\}$, mätt i timmar. För att kretsen ska fungera behöver minst två av transistorerna vara hela. Vi vill att kretsen ska fungera minst 4000 timmar.

(a) Beräkna väntevärde och median för livslängden av den första transistorn. [2p]

(b) Beräkna sannolikheten att transistor 1 går sönder inom 4000 timmar. [2p]

(c) Beräkna sannolikheten att kretsen fungerar minst 4000 timmar. [3p]

(d) Efter 2000 timmar undersöks transistorerna och två av dem fungerar fortfarande. Beräkna sannolikheten att kretsen fungerar efter ytterligare 2000 timmar, det vill säga efter totalt 4000 timmar. [3p]

Lösning:

(a) Väntevärdet ges av

$$\mathbf{E}(X_1) = \frac{1}{\lambda} = \frac{1}{10^{-4}} = 10^4.$$

För medianen vill vi lösa ekvationen $F_{X_1}(x_{0.5}) = 0.5$, det vill säga

$$\frac{1}{2} = 1 - \mathrm{e}^{-10^{-4} x_{0.5}},$$

vilket ger $x_{0.5} = 10^4 \ln(2) = 6931$.

(b) Vi har

$$\mathbf{P}(X_1 < 4000) = F_{X_1}(4000) = 1 - \mathrm{e}^{-0.4} = 0.33.$$

(c) Låt Z vara antalet transistorer som är trasiga efter 4000 timmar, och låt $p = 1 - \mathbf{P}(X_1 < 4000) = 0.67$ vara sannolikheten att en given transistor är hel efter 4000 timmar. Vi har

$$\begin{aligned}\mathbf{P}(Z \leq 1) &= \mathbf{P}(Z = 0) + \mathbf{P}(Z = 1) \\ &= p^3 + 3p^2(1-p) \\ &= 0.67^3 + 3 \cdot 0.67^2 \cdot 0.33 \\ &= 0.30 + 0.44 = 0.74.\end{aligned}$$

(d) Exponentialfördelade slumpvariabler saknar minne, så detta är samma sak som att fråga om två transistorer båda är hela efter 2000 timmar. Detta ges av

$$(1 - F_{X_1}(2000))^2 = \mathrm{e}^{-0.4} = 0.67.$$

4. Vid Örebro universitet arbetar 1000 lärare. IT-avdelningen behöver avgöra hur väl lärarnas datorer fungerar, och kontaktar därför 20 slumpmässigt utvalda lärare för för att fråga om de har möjlighet att lämna ifrån sig sin dator under mellandagarna för en bedömning av datorernas skick. Hundra av lärarna ägnar mellandagarna åt att konstruera tentor och har inte möjlighet att lämna ifrån sig sin dator, medan övriga lärare inte har något emot att lämna ifrån sig datorn.

   (a) Beräkna, med lämplig approximation, sannolikheten att exakt 17 lärare lämnar ifrån sig sin dator. [5p]

   (b) Redan efter en dag inser IT-avdelningen att de behöver göra en bredare undersökning. De kontaktar därför ytterligare 180 lärare med samma fråga. Beräkna, med lämplig approximation, sannolikheten att de från dessa totalt 200 lärare får in minst 175 datorer. [5p]

Lösning:

(a) Antalet lärare som inte lämnar ifrån sig en dator är $X \sim \text{Hyp}(1000, 20, 0.1)$. Vi kan inte beräkna $\mathbf{P}(X = 3)$ direkt, men eftersom $n/N = 2$ % kan vi approximera med en binomialfördelad slumpvariable, närmare bestämt $Y \sim \text{Bin}(20, 0.1)$. Enligt denna fördelning får vi, med hjälp av tabell 2,

$$\mathbf{P}(X = 3) \approx \mathbf{P}(Y = 3) = F_Y(3) - F_Y(2) = 0.8670 - 0.6769 = 0.1901.$$

(b) Låt de lärare som lämnar ifrån sig sina datorer ges av $Z \sim \text{Hyp}(1000, 200, 0.1)$. Med 200 tillfrågade lärare kan vi inte längre approximera med binomialfördelning, och inte heller Poisson är möjlig, men eftersom

$$V(Z) = np(1-p)\frac{N-n}{N-1} = 200 \cdot 0.1 \cdot 0.9 \frac{1000-200}{1000-1} = 18\frac{800}{999} = \frac{1600}{111} > 5$$

kan vi approximera med en normalfördelad slumpvariabel $U \sim N(20, 1600/111)$. Låt vidare $V = \frac{Y-20}{40/\sqrt{111}} \sim N(0,1)$. Med halvkorrektion fås

$$\begin{aligned}\mathbf{P}(Z \leq 25) &\approx \mathbf{P}(U \leq 25.5)\\ &= \mathbf{P}\left(V \leq \frac{25.5-20}{40/\sqrt{111}}\right)\\ &= \Phi\left(\frac{5.5}{40/\sqrt{111}}\right)\\ &= \Phi(1.45) = 0.9265.\end{aligned}$$

5. Betrakta täthetsfunktionen $f_X(x) = \theta(1+x)^{-(\theta+1)}$, där $x \geq 0$. Vi vill skatta parametern θ. Till vår hjälp av vi observationerna $x_1 = 0.4$ och $x_2 = 0.8$.

(a) Man kan visa att $\mathbf{E}(X) = \frac{1}{\theta-1}$. Använd momentmetoden för att skatta θ utgående från dessa observationer. [5p]

(b) Vi får nu information om att $\theta \in \{2, 3, 4\}$. Beräkna likelihoodfunktionen för dessa tre värden och bestäm därefter ML-skattningen av θ. [5p]

Lösning:

(a) Vi ställer upp ekvationen

$$\frac{1}{\theta^*-1} = \mathbf{E}(X) = m(\theta^*) = \bar{x} = \frac{0.4+0.8}{2} = \frac{1.2}{2} = \frac{3}{5},$$

som ger $\theta^* - 1 = 5/3$ eller $\theta^* = 8/3$.

(b) Vi beräknar $L(\theta)$ för dessa tre möjliga värden på θ:

$$L(2) = 2(1+0.4)^{-3} \cdot 2(1+0.8)^{-3} = 0.250$$
$$L(3) = 3(1+0.4)^{-4} \cdot 3(1+0.8)^{-4} = 0.223$$
$$L(4) = 4(1+0.4)^{-5} \cdot 4(1+0.8)^{-5} = 0.157$$

Av dessa ger $\theta = 2$ det högsta värdet, så vår skattning blir $\theta^* = 2$.

6. Vid provvägning av granolaförpackningar med angiven vikt 400 gram fås följande resultat: [10p]

408 387 367 421 396 403 366 384 405 363

Innehåller dessa paket tillräckligt med granola? Vid en närmare kontroll har tillverkaren angivit att paketens vikt ska ha fördelningen $N(400, 400)$. Under antagandet att standardavvikelsen alltjämt är 20 gram, pröva hypotesen $H_0 : \mu = 400$ mot lämplig mothypotes på nivån 5%.

Lösning: Eftersom vi enbart är oroliga att förpackningarna väger för litet så blir mothypotesen $H_1 : \mu < 400$. Vi väljer därför att beräkna ett konfidensintervall på formen

$$I_\mu = (0, \bar{x} + \lambda_\alpha \frac{\sigma}{\sqrt{n}}),$$

där $\alpha = 0.05$, $\sigma = 20$ och $n = 10$, och avgör sedan frågan genom att se om intervallet innehåller talet 400. Tabell 5 ger $\lambda_{0.05} = 1.645$ och summering av våra mätvärden ger $\bar{x} = 3900/10 = 390$. Vi får därmed intervallet

$$I_\mu = (0, 390 + 1.645 \frac{20}{\sqrt{10}}) = (0, 390 + 10.4) = (0, 400.4).$$

Eftersom 400 ingår i detta intervall kan vi inte förkasta nollhypotesen, och kan därför inte invända mot den som anser att paketen innehåller tillräckligt med granola.